

Normalized Entropy Vectors, Network Information Theory and Convex Optimization

Babak Hassibi

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
hassibi@caltech.edu

Sormeh Shadbakht

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
sormeh@caltech.edu

Abstract—We introduce the notion of normalized entropic vectors—slightly different from the standard definition in the literature in that we normalize entropy by the logarithm of the alphabet size. We argue that this definition is more natural for determining the capacity region of networks and, in particular, that it smooths out the irregularities of the space of non-normalized entropy vectors and renders the closure of the resulting space convex (and compact). Furthermore, the closure of the space remains convex even under constraints imposed by memoryless channels internal to the network. It therefore follows that, for a large class of acyclic memoryless networks, the capacity region for an arbitrary set of sources and destinations can be found by maximization of a linear function over the *convex* set of channel-constrained normalized entropic vectors and some linear constraints. While this may not necessarily make the problem simpler, it certainly circumvents the “infinite-letter characterization” issue, as well as the nonconvexity of earlier formulations, and exposes the core of the problem. We show that the approach allows one to obtain the classical cutset bounds via a duality argument. Furthermore, the approach readily shows that, for acyclic memoryless *wired* networks, one need only consider the space of unconstrained normalized entropic vectors, thus separating channel and network coding—a result very recently recognized in the literature.

I. INTRODUCTION

In recent years, there has been a growing interest in information transmission over networks. While, historically, information theory has played a central role in the development of point-to-point communication systems, it is fair to say that it has had a far less impact on the design of most of the networks currently in use, and especially the Internet. Information-theoretic considerations have had a minimal impact on the development of protocols (such as UDP or TCP-IP), routing, congestion control, etc. The absence of an interaction between networking and information theory, until very recently, can be attributed to many facts. The most important, however, must be the fact that, unlike point-to-point communication problems where the information-theoretic capacity has long been known, the capacity region of almost all *multi-user* information theory problems is open. The handful, or so, of cases where the capacity region is known happen to be where certain proposed schemes have been fortunate enough to coincide with the (rather obvious) cutset outer bounds. In any event, anything close to a “theory” of multi-user information theory is still lacking.

While, “in principle”, it is possible to write down a characterization for the capacity region of most network information theory problems, the difficulty is that this characterization is *infinite-letter* and *non-convex*. In other words, evaluating the capacity region requires solving an infinite succession of non-convex optimization problems over certain distributions whose number of variables goes to infinity. This is in stark contrast with point-to-point (single-user) memoryless channels where the characterization is both *single-letter* and *convex*.

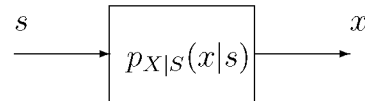


Fig. 1. A point-to-point communication problem.

To make this more explicit, consider the point-to-point memoryless channel of Fig. 1. The capacity is clearly

$$C = \max_{p_S(\cdot)} I(S; X) = \max_{p_S(\cdot)} \{H(X) - H(X|S)\}, \quad (1)$$

where $p_S(\cdot)$ is the input distribution and $H(X)$ and $H(X|S) = H(X, S) - H(S)$ are the usual entropy and conditional entropies. Problem (1) is referred to as single letter, since all entropies are over only a single channel use. The problem is one of “convex optimization” since $I(X; Y)$ is a concave function of the input distribution and so we are maximizing a concave function.

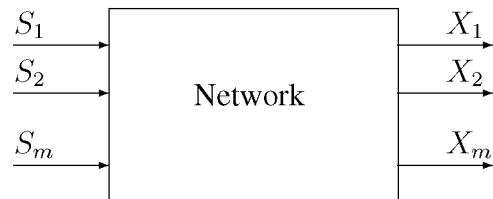


Fig. 2. A communication problem over an acyclic memoryless network.

Consider now the network problem of Figure 2. Assume that the network is acyclic and memoryless (in the sense that all channels internal to the network are memoryless) and that there is no feedback from the destinations to the

sources. Suppose that each source S_i needs to transmit to its corresponding destination X_i at some rate R_i , $i = 1, \dots, m$.¹ Now in this case it is not too hard to show that the *rate region*, defined as the set of rates $\{R_i\}_{i=1}^m$ which can be reliably exchanged between the sources and destinations, is given by (see, e.g., [1], [2], [3]):

$$\mathcal{R} = \text{cl} \left\{ R_i, i = 1, \dots, m \mid R_i < \frac{1}{T} (H(X_i^T) - H(X_i^T | S_i^T)) \right\} \quad (2)$$

as $T \rightarrow \infty$

where $\text{cl}\{\cdot\}$ refers to the closure of the set. Here S_i^T and X_i^T are random variables obtained from concatenating the corresponding source and destination random variables over T channel uses. Of course, the characterization of the rate region \mathcal{R} as in (2) is not surprising—in some sense it can be considered almost as the definition of the rate region. Computing it, however, is another matter.

An equivalent way of representing the rate region is through its tangent hyperplanes. These can be obtained via solving the following optimization problem

$$\lim_{T \rightarrow \infty} \sup_{\substack{p_{S_i^T}(\cdot) \text{ and} \\ \text{network operations}}} \sum_{i=1}^m \alpha_i \frac{1}{T} (H(X_i^T) - H(X_i^T | S_i^T)) \quad (3)$$

where $\{\alpha_i\}_{i=1}^m$ represents the normal vector to the tangent hyperplane, and where “network operations” represents all permissible internal operations of the network. The problem (3) is notoriously difficult since it is *infinite-letter* (i.e., it involves optimization over joint distributions whose number of variables goes to infinity) and *non-convex* (the cost function $\sum_{i=1}^m \alpha_i \frac{1}{T} (H(X_i^T) - H(X_i^T | S_i^T))$ is highly non-convex in the $p_{S_i^T}(\cdot)$ and “network operations”). For this reason, the characterization of (3) has very rarely been explicitly used.²

The goal of this paper is to suggest an alternative representation and study of the aforementioned network information theory problem. The main idea is to define the space of (suitably normalized) entropic vectors and to show that a very wide range of network information theory problems reduce to the optimization of a linear cost over the convex set of (constrained) entropic vectors. This viewpoint has several advantages: first, it does away with the complications of infinite-letter characterizations (in fact, the infinite limit simplifies the representation considerably), second, it renders the problem convex and, third, it shows how through duality one may obtain classical results such as cutset bounds. While by no means solving the network information theory problem in itself, it does point to what the heart of the problem is:

¹We should note that this formulation can allow for quite general problems. If a source is desired by many destinations, then all we need to do is repeat that source as many times as desired. If, on the other hand, a destination requires many sources, all we need to do is to repeat that destination as many times as necessary.

²The only work that we are aware of that uses the infinite-letter characterization (2) is [4], which shows that it can be reduced to a single letter characterization for memoryless multiple-access channels.

characterizing the space of (channel constrained) entropic vectors.

The next section defines the notion of entropic vectors and shows that the resulting space is convex. This is then used to formulate network information theory problems as convex optimizations. In section III cutset bounds, as well as a separation between network coding and channel coding are studied as some special instances of this formulation.

II. NETWORK INFORMATION THEORY AND ENTROPY VECTORS

A. Entropy Region

Recall that the objective (3) is just a linear function of entropies:

$$\sum_{i=1}^m \alpha_i \frac{1}{T} (H(X_i^T) + H(S_i^T) - H(X_i^T, S_i^T)). \quad (4)$$

This motivates the following definition. Let X_1, \dots, X_n be a collection of n jointly distributed discrete random variables with alphabet size N each.³ For any set $\mathcal{S} \subseteq \{1, \dots, n\}$, we define the *normalized entropy* as

$$h_{\mathcal{S}} = \frac{1}{\log N} H(X_i, i \in \mathcal{S}). \quad (5)$$

We remark that the motivation for the definition of the normalized entropy is the fact that what appears in (4), i.e., $\frac{1}{T} H(X_i^T)$, $\frac{1}{T} H(S_i^T)$ and $\frac{1}{T} H(X_i^T, S_i^T)$ are essentially normalized entropies since the alphabet-sizes of S_i^T and X_i^T are exponential in T . Now there exist $2^n - 1$ non-empty such sets \mathcal{S} . The $2^n - 1$ dimensional vector obtained from stacking these entropies in a column vector, say, is called an *entropy vector*.

Definition 1 (Entropic Vectors): Any vector $v \in \mathcal{R}^{2^n - 1}$ which can be regarded as the entropy vector of some collection of n random variables, for some value of N , is called entropic. We will denote the space of entropic vectors by Ω_n^* .

We should remark that our definition of entropic vectors is slightly different from what is conventionally used in the literature (see, e.g., [5]). The convention is to consider entropy vectors without the normalization. In this case, the space of entropy vectors is referred to as Γ_n^* . We believe our definition to be more natural. One indication is the more direct connection to (3) and (4). The other is the fact that the set Γ_n^* is quite complicated: it has an irregular boundary [6] and many “holes”. Its closure, $\bar{\Gamma}_n^*$, is therefore more often studied, which can be shown to be a convex cone (see, e.g., [5], [7]). The set Ω_n^* is, however, much simpler. It is clearly bounded, since

$$h_{\mathcal{S}} \leq |\mathcal{S}|, \quad (6)$$

where $|\mathcal{S}|$ is the cardinality of the set \mathcal{S} . Furthermore, it is straightforward to show that the closure of Ω_n^* is a convex set.

³There is no loss of generality in this assumption. If the random variables have different alphabet-sizes, we can always take N to be the largest alphabet-size and to make the probability mass functions zero wherever appropriate.

Theorem 1 (Convexity of $\bar{\Omega}_n^$):* The closure of the set of entropic vectors, $\bar{\Omega}_n^*$ is convex.

We remark that, for any fixed N , the set of entropic vectors is highly non-convex. It is the fact that N is arbitrary (and can grow unbounded) that yields convexity. We will present two proofs, since both are instructive.

Proof 1: (Time sharing) Suppose $h_x \in \bar{\Omega}_n^*$, corresponding to random variables X_1, \dots, X_n with alphabet-size N_x and $h_y \in \bar{\Omega}_n^*$ corresponding to random variables Y_1, \dots, Y_n with alphabet-size N_y . Make n_x independent copies of the first set and n_y independent copies of the second so that together the new concatenated random variables have alphabet-size $N_x^{n_x} N_y^{n_y}$. The resulting entropy vector is clearly

$$\frac{n_x \log N_x}{n_x \log N_x + n_y \log N_y} h_x + \frac{n_y \log N_y}{n_x \log N_x + n_y \log N_y} h_y,$$

which, since n_x and n_y are arbitrary, implies that one can get arbitrarily close to any point on the convex hull of h_x and h_y . This implies the convexity of the closure of Ω_n^* . ■

Now if we form the convex combination of the distributions of two sets of random variables X_i and Y_i , $i = 1, \dots, n$, with fixed alphabet-size N :

$$pZ_1, \dots, Z_n(z_1, \dots, z_n) = p\theta p_{X_1, \dots, X_n}(z_1, \dots, z_n) + (1 - p\theta) p_{Y_1, \dots, Y_n}(z_1, \dots, z_n) \quad (7)$$

to obtain the set of random variables Z_i , it is certainly *not* true that

$$h_z = p\theta h_x + (1 - p\theta) h_y. \quad (8)$$

However, the next proof shows that this is true in the limit!

Proof 2: (Convex combination of distributions) Make T independent copies of each of the sets of random variables X_i and Y_i and consider the distribution

$$p\theta \prod_{t=1}^T p_{X_1, \dots, X_n}(z_1^t, \dots, z_n^t) + (1 - p\theta) \prod_{t=1}^T p_{Y_1, \dots, Y_n}(z_1^t, \dots, z_n^t) \quad (9)$$

Now for any $\mathcal{S} \subseteq \{1, \dots, n\}$, we have

$$\frac{H(Z_{\mathcal{S}}^T | \theta)}{p\theta H(X_{\mathcal{S}}^T) + (1 - p\theta) H(Y_{\mathcal{S}}^T)} \leq H(Z_{\mathcal{S}}^T) \leq \frac{H(Z_{\mathcal{S}}^T, \theta)}{= H(Z_{\mathcal{S}}^T | \theta) + H(p\theta)} \quad (10)$$

Normalizing by $\log N^T$ yields

$$p\theta h_x + (1 - p\theta) h_y \leq h_z \leq p\theta h_x + (1 - p\theta) h_y + \frac{-p\theta \log p\theta - (1 - p\theta) \log(1 - p\theta)}{T \log N} \quad (11)$$

which shows the convexity of the closure as $T \rightarrow \infty$. ■

We end this section by emphasizing that our choice of normalized entropy vectors, and letting N be arbitrary, retains all the information needed to solve network information theory problems, yet “smooths out” all the irregularities in Γ_n^* . In fact, the relationship between the two sets is as follows:

Theorem 2 (Ω_n^ and Γ_n^*):* Define the ray of a set \mathcal{S} as

$$\text{ray}(\mathcal{S}) = \{\alpha X | \alpha > 0, X \in \mathcal{S}\}. \quad (12)$$

Then we have

$$\text{ray}(\bar{\Omega}_n^*) = \bar{\Gamma}_n^*, \quad (13)$$

i.e., the ray of $\bar{\Omega}_n^*$ is the closure of Γ_n^* .

Proof: Let $V \in \bar{\Gamma}_n^*$. This means that for any $\epsilon > 0$ and $\mathcal{S} \subset \{1, \dots, n\}$ there exists random variables X_1, \dots, X_n of some alphabet size N such that $|H(X_{\mathcal{S}}) - V_{\mathcal{S}}| \leq \epsilon$. Therefore $\frac{1}{\log N} V \in \bar{\Omega}_n^*$ and so $V \in \text{ray}(\bar{\Omega}_n^*)$.

Conversely suppose $V \in \text{ray}(\bar{\Omega}_n^*)$; then by definition there exists an α such that $\frac{1}{\alpha} V \in \bar{\Omega}_n^*$, from which it follows that for any $\epsilon \geq 0$ there exist random variables X_1, \dots, X_n with alphabet size N for which $|h(X_{\mathcal{S}}) - \frac{1}{\alpha} V_{\mathcal{S}}| \leq \epsilon$. Thus, $\frac{\log N}{\alpha} V$ is a non-normalized entropic vector. Since Γ_n^* is a convex cone [5], this implies that $V \in \bar{\Gamma}_n^*$. ■

B. Network Information Theory

Let us now return to the network problem (3) and study the consequences of what we have considered so far. Consider *all* the random variables in the network and designate them by X_i , $i = 1, \dots, n$ (the X_i will thus include both the sources, the destinations, as well as any random variables internal to the network). Now due to the normalization $\frac{1}{T}$ in (3), we can simply write the objective as a linear combination of entropic vectors constructed from the X_i . Furthermore, since we consider the closure of the set of entropic vectors, the $\lim_{T \rightarrow \infty}$ does not pose any problems. Finally, since the set of entropic vectors is dense in its closure, replacing optimization over Ω_n^* by optimization over $\bar{\Omega}_n^*$ does not cause a problem.

The upshot of all these arguments is that (3) can be rewritten as

$$\sup \alpha^T h, \quad (14)$$

(where \cdot^T now refers to transpose) subject to $h \in \bar{\Omega}_n^*$ and subject to the constraints imposed by the network. These are of two kinds.

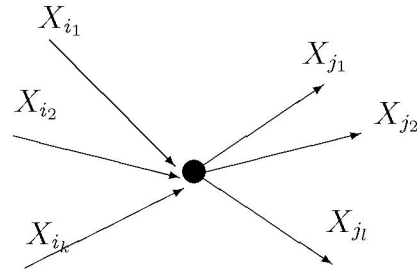


Fig. 3. Topological constraints at any non-source node.

1) Topological Constraints: Topological constraints have to do with the information flow in the network. Thus, consider a “non-source” node in the network with incoming messages $\{X_{i_p}\}_{p=1}^k$ and outgoing messages $\{X_{j_q}\}_{q=1}^l$ (see Fig. 3). Then clearly, we have the following linear constraints on the entropy

$$h(X_{j_q} | X_{i_1}, \dots, X_{i_k}) = 0 \quad (15)$$

or, equivalently:

$$h(X_{j_q}, X_{i_1}, \dots, X_{i_k}) - h(X_{i_1}, \dots, X_{i_k}) = 0 \quad (16)$$

for all $q = 1, \dots, l$. Alternatively, at source nodes we have $h(S_i, S_j) - h(S_i) - h(S_j) = 0$, if source nodes i and j are independent or $h(S_i, S_j) = h(S_i) = h(S_j)$, if source nodes i and j are identical.

The conclusion is that topological constraints simply introduce linear constraints on the entries of the entropy vector.

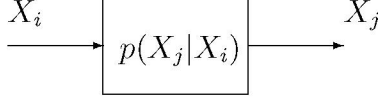


Fig. 4. A channel internal to the network.

2) *Channel Constraints*: Channel constraints do not translate directly to entropies. What they do is constrain the joint distribution of all random variables in the network (which then determines the admissible entropy vectors). Thus, referring to Fig. 4 let a certain discrete memoryless channel relate the messages X_i and X_j . Thus,

$$p(X_i, X_j) = p(X_j|X_i)p(X_i), \quad (17)$$

or, equivalently,

$$\begin{aligned} & \int \prod_{k \neq i, j} dX_k p(X_1, \dots, X_n) \\ &= p(X_j|X_i) \int \prod_{k \neq i} dX_k p(X_1, \dots, X_n) \end{aligned} \quad (18)$$

which is simply a linear constraint on the joint distribution. Now the fact that the underlying distribution satisfies linear constraints has no effect on the validity of the two proofs we gave for Theorem 1. Therefore the presence of channels inside the network does not affect the convexity of the space of admissible entropy vectors. We formalize this result in the following theorem.

Theorem 3 (Channel-constrained entropic vectors): Let $\Omega_{n,c}^*$ denote the space of entropic vectors that are constrained by the discrete memoryless channels in the network. Then the closure of this set, i.e., $\bar{\Omega}_{n,c}^*$, is convex.

C. Network Information Theory as Convex Optimization

From the above discussions we can conclude that the problem (3) is equivalent to

$$\max_{h \in \bar{\Omega}_{n,c}^*, Ah=0} \alpha^T h, \quad (19)$$

where $\bar{\Omega}_{n,c}^*$ denotes the convex space of channel-constrained entropic vectors and $Ah = 0$ represents the topological constraints. Note that, since the constraint set is closed, we can use max, rather than sup.

Remarks: The formulation (19) is significant for at least two reasons.

- 1) By going to the space of normalized entropy vectors, we have circumvented the problem of “infinite-letter characterization”.

- 2) We have also circumvented the “non-convexity”. (19) is a convex optimization problem. In fact, the infinite-letter characterization is what yields convexity (the space of entropic vectors is not convex for any finite T).

III. SOME APPLICATIONS

A. Duality and Cutset Bounds

As a first attempt, a simple use of some basic machinery from convex optimization yields some interesting results. In network flow problems, the duality between *max-flow* and *min-cut* is well known [8], [9]. In information theory cutset outerbounds are also well known (see, e.g., [10]); however, to the best of our knowledge, these have been obtained by relaxing the network problem to a point-to-point problem (assuming certain nodes can fully cooperate with the sources and others can fully cooperate with the destinations), rather than through any duality argument.

Note that in (19) we can enforce the linear constraints through a Lagrange multiplier λ to obtain

$$\max_{h \in \bar{\Omega}_{n,c}^*, Ah=0} \alpha^T h = \max_{h \in \bar{\Omega}_{n,c}^*} \min_{\lambda} (\alpha^T h + \lambda^T Ah). \quad (20)$$

Using the duality of convex optimization we can interchange the max and min to obtain

$$\max_{h \in \bar{\Omega}_{n,c}^*, Ah=0} \alpha^T h = \min_{\lambda} \max_{h \in \bar{\Omega}_{n,c}^*} (\alpha^T h + \lambda^T Ah). \quad (21)$$

In particular, for any λ , we have the upper bound

$$\max_{h \in \bar{\Omega}_{n,c}^*, Ah=0} \alpha^T h \leq \max_{h \in \bar{\Omega}_{n,c}^*} (\alpha^T h + \lambda^T Ah). \quad (22)$$

Consider now an arbitrary cut through the network, such that all the source nodes reside on one side of the cut and all the destination nodes on the other side of the cut. Set to zero all components of the Lagrange multiplier λ that correspond to edges that do not cross the cut. Then the upper bound in (22) after minimizing over the remaining components of λ , is simply the cutset upper bound corresponding to this cut.⁴

We have therefore obtained an interpretation of cutset bounds through duality and Lagrange multipliers. More clever choices of the Lagrange multiplier can lead to improved upper bounds over the cutset bound.

B. Wired Networks

In the current framework, solving network information theory problems requires characterizing the set $\bar{\Omega}_{n,c}^*$. This seems formidable (to say the least). However, as we shall presently see, for wired networks things simplify considerably. Wired networks are defined through three main characteristics.

- 1) Each link represents a (discrete memoryless) channel.
- 2) The signals transmitted on outgoing edges of a node (X_i, X_j in Fig. 5, say) can be distinct.
- 3) The signals impinging on a node (X_k, X_l in Fig. 5, say) are received without interference.

⁴We omit the details for lack of space; the basic idea is that once the components of the Lagrange multiplier that do not cross the cut are set to zero all nodes on either side of the cut can fully cooperate and so the problem becomes a point-to-point problem whose value is simply the cut capacity.

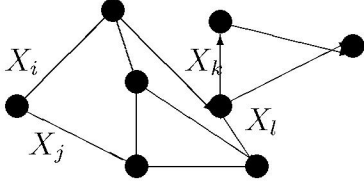


Fig. 5. A wired network.

In a wired network with k edges we therefore have $2k$ random variables: for each edge i a transmitted message X_i (at the input of the channel) and a corresponding received message Y_i (at the output of the channel). If the channel has scaled capacity C_i (i.e., scaled by the \log of the maximum of the input and output alphabet sizes of the channel) it clearly imposes the constraint

$$h(X_i) + h(Y_i) - h(X_i, Y_i) \leq C_i, \quad i = 1 \dots, k. \quad (23)$$

We can therefore define random variables Z_i , such that

$$h(Z_i) = h(X_i) + h(Y_i) - h(X_i, Y_i) \leq C_i \quad i = 1 \dots, k. \quad (24)$$

Furthermore, we will assume Z_i to have the interpretation of being the information that can be transmitted through the channel error-free. We can assume that X_i , Z_i and Y_i satisfy the Markov chain

$$X_i \rightarrow Z_i \rightarrow Y_i. \quad (25)$$

The fact that Z_i can be recovered perfectly from Y_i can be expressed as

$$h(Z_i|Y_i) = 0. \quad (26)$$

And by symmetry of the mutual information with respect to X_i and Y_i we can assume:

$$h(Z_i|X_i) = 0. \quad (27)$$

If we denote the space of admissible entropic vectors for the Z_i by $\bar{\Omega}_{k,c,z}^*$, it is easy to see that

$$\bar{\Omega}_{k,c,z}^* = \bar{\Omega}_k^* \cap \{h \mid h_{1:k} \leq C\}, \quad (28)$$

where C is the vector of scaled link capacities. In other words, the space of entropic vectors for the Z_i is simply the space of entropic vectors where each $h(Z_i)$ is constrained by the link capacity in (24).

In the wired case, since the signals on the different edges are not identical and do not interact (through interference, or otherwise), it turns out that all the quantities of interest can be expressed in terms of the Z_i . In fact, going back to Figure 2 and formulas (2) and (3), we see that these quantities are the mutual information between every source-destination pair, i.e., $I(X_i, S_i)$. Thus, denoting by Z_{S_i} the error-free signal on the source link i and by Z_{X_i} the error-free signal on the destination link i , we have the Markov chain

$$S_i \rightarrow Z_{S_i} \rightarrow Z_{X_i} \rightarrow X_i. \quad (29)$$

We can now show

$$\begin{aligned} I(S_i; X_i) &= h(S_i) - h(S_i|X_i) \\ &= h(S_i) - h(S_i, Z_{X_i}|X_i) \quad \text{by (26)} \\ &= h(S_i) - \underbrace{h(Z_{X_i}|X_i)}_{=0} - h(S_i|X_i, Z_{X_i}) \\ &= h(S_i) - h(S_i|Z_{X_i}) \quad \text{by (29)} \\ &= h(Z_{X_i}) - h(Z_{X_i}|S_i) \\ &= h(Z_{X_i}) - h(Z_{X_i}, Z_{S_i}|S_i) \quad \text{by (27)} \\ &= h(Z_{X_i}) - \underbrace{h(Z_{S_i}|S_i)}_{=0} - h(Z_{X_i}|Z_{S_i}, S_i) \\ &= h(Z_{X_i}) - h(Z_{X_i}|Z_{S_i}) \quad \text{by (29)} \\ &= I(Z_{S_i}; Z_{X_i}), \end{aligned}$$

which implies that the objective function in (3) can be expressed solely in terms of the Z_i . From the above argument we can easily conclude the following theorem. (This result has already been observed by other authors [11].)

Theorem 4 (Separation of Channel and Network Coding): Consider an acyclic memoryless wired network with k edges and assume that there is no feedback from destinations to sources. Then any network information theory problem reduces to a problem of the form

$$\max_{h \in \bar{\Omega}_k^*, h_{1:k} \leq C, Ah=0} \alpha^T h, \quad (30)$$

where $Ah = 0$ represents the topological constraints of the network and C is the vector of edge capacities.

This result is significant for two reasons.

- 1) For wired networks, the channels affect the rate region only through their capacities.
- 2) For wired networks, determining the rate region requires determining only $\bar{\Omega}_k^*$.

REFERENCES

- [1] C.E. Shannon, "Two-way communications channels," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1961, pp. 611–644.
- [2] R. Ahlswede, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. 23, pp. 1–37, January 1977.
- [3] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, pp. 4–21, January 2003.
- [4] R.S. Cheng and S. Verdú, "On limiting characterizations of memoryless multiuser capacity regions," *IEEE Transactions on Information Theory*, vol. 39, pp. 609–612, March 1993.
- [5] R.W. Yeung, *A first course in information theory*, Kluwer, 2002.
- [6] F. Matus, "Infinitely many information inequalities," in *Submitted to ISIT*, 2007.
- [7] Zhen Zhang and Raymond Yeung, "A non-shannon-type conditional inequality of information quantities," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1982–1986, 1997.
- [8] P. Elia, A. Feinstein, and C.E. Shannon, "Note on maximum flow through a network," *IRE Transactions on Information Theory*, vol. 2, pp. 117–119, 1956.
- [9] L.R. Ford and D.R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.
- [10] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
- [11] R. Koetter, "Separation for source, channel and network coding," in *IEEE Information Theory Workshop—Keynote Speech*, 2006.